

УДК 681.327.12

М.М.Биков, к.т.н., проф.; В.В.Ковтун, к.т.н., доц.; А. Раїмі, PhD, доц.**МЕТОД ОПТИМІЗАЦІЇ ПРОЦЕСУ НАВЧАННЯ НЕЙРОМЕРЕЖІ В
ЗАДАЧІ РОЗПІЗНАВАННЯ МОВЦІВ**

Авторами досліджено вплив нейромережевого класифікатора на ефективність роботи системи розпізнавання мовців, проведено моделювання таких систем та досліджено вплив параметрів класифікатору на точність розпізнавання.

Ключові слова: ідентифікація мовця, розпізнавання голосів, класифікація, прийняття рішень, нейромережа.

Вступ

Процес ідентифікації особи за індивідуальними особливостями голосу можна розділений на два основних етапи, а саме, отримання оцифрованого запису мовного сигналу (парольної фрази) та декодування.

На першому етапі вхідний аудіо сигнал записується і проходить попереднє оброблювання, яке, зазвичай, включає в себе фільтрацію, сегментацію на нормативні для розпізнавання відрізки та нормалізацію тривалості звучання виділених відрізків-фреймів. на фрагменти (сегментується). На етапі декодування отриманий цифровий сигнал аналізується з використанням різноманітних методів для отримання образу, що відповідає аналізованому сигналу, у просторі інформативних для розпізнавання особи мовця ознак із подальшим віднесенням цього образу до одного із сформованих в просторі ознак кластерів, які відповідають особам мовців, на розпізнавання яких систему навчено. Цю функцію виконує класифікатор (найчастіше це нейронна мережа – трьохшаровий персептрон [1-3]), коректна робота якого, в решті решт, визначає достовірність роботи системи в цілому. Отже, задача формування, навчання та оптимізації нейромережевого класифікатора для розпізнавання мовців є актуальною.

Одним із популярних середовищ для цифрового аналізу є крос-платформна бібліотека ALGLIB (www.alglib.net), яка є сумісною із такими сучасними мовами програмування як C#, C++, VBS в операційних системах Windows, Linux, Solaris. В цій бібліотеці реалізовано, зокрема, і ряд алгоритмів цифрового аналізу та обробки сигналів, оптимізації, створення та налаштування нейромереж тощо. Так нейронна мережа в ALGLIB представлена структурою MultiLayerPerceptron. Робота з нейронними мережами передбачає послідовне виконання таких операцій, як вибір архітектури та ініціалізація структури за допомогою відповідної функції; навчання нейронної мережі за допомогою одного з алгоритмів; використання навченої мережі.

Авторами було створено систему розпізнавання мовців із наведеною на рис. 1 структурою.

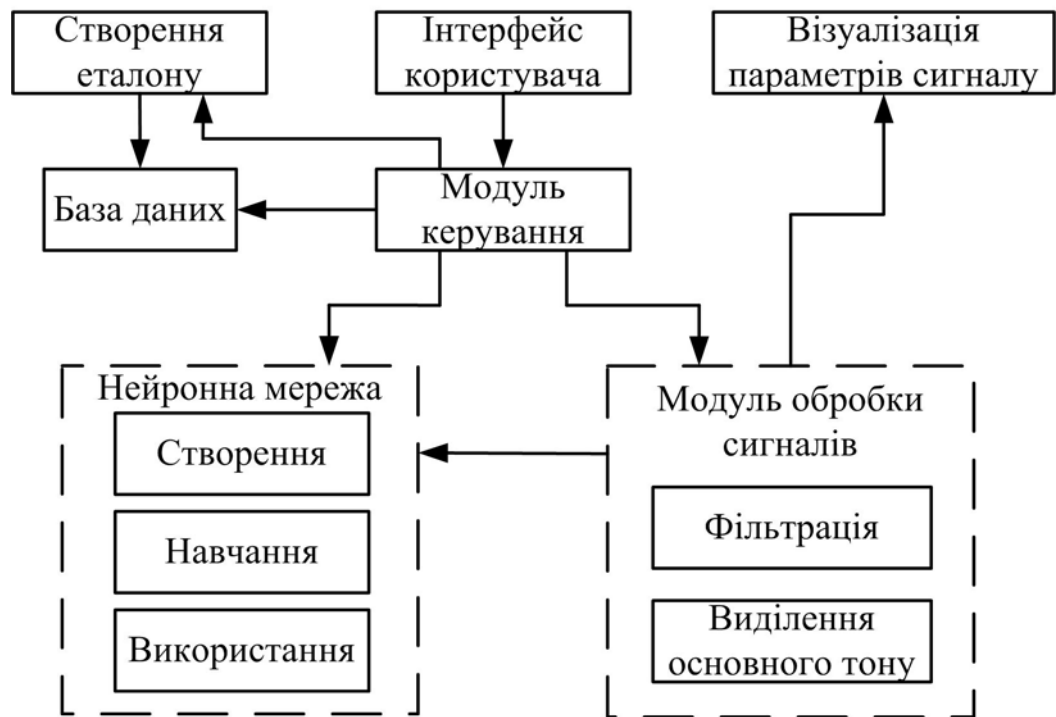


Рисунок 2.1 – Структура системи розпізнавання мовців

Моделювання системи розпізнавання мовців засобами UML

Моделювання системи засобами UML дозволяє всебічно та доступно представити модель створюваної системи, зокрема, системи розпізнавання мовців. Для опису системи було побудовано діаграму варіантів використання - вихідне концептуальне уявлення або концептуальна модель системи в процесі розробки і впровадження, представлену на рис. 2.



Рисунок 2 – UML діаграма варіантів використання

На рисунку 3 представлено діаграму діяльності. Кожен стан на діаграмі діяльності відповідає виконанню певної операції, а перехід в наступний стан відбувається тільки після завершення виконання цієї операції. Діаграма діяльності подано у формі графа діяльності, вершинами якого є стани дії або діяльності, а дугами - переходи від одного стану дії до іншого.

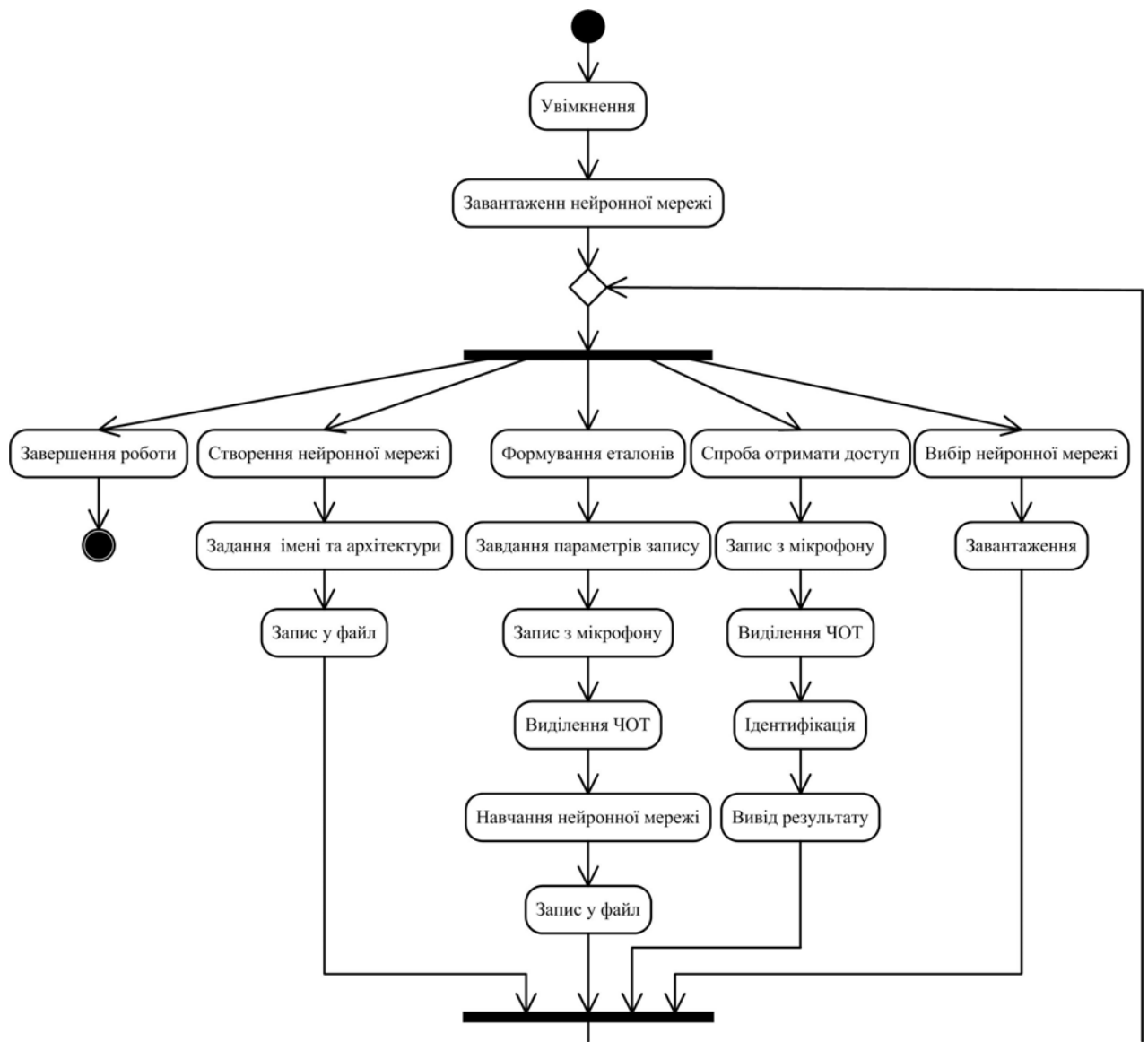


Рисунок 3 – UML діаграма діяльності

Діаграма класів (class diagram) призначена для представлення статичної структури моделі системи в термінології класів об'єктно-орієнтованого програмування. На цій діаграмі показано класи, інтерфейси, об'єкти та кооперації, а також їхні відносини. На рисунку 4 представлено діаграму класів системи розпізнавання мовців.

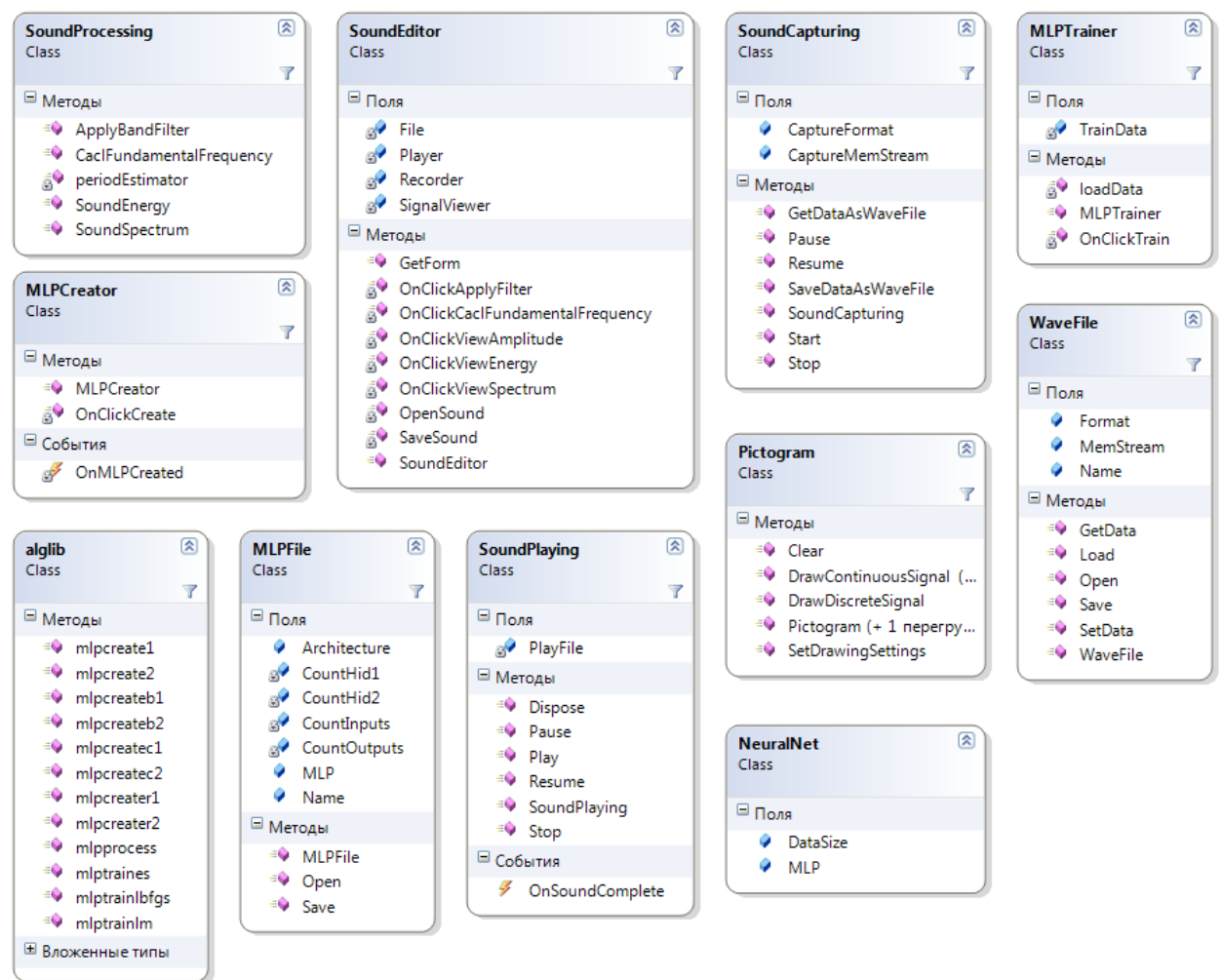


Рисунок 4 – Діаграма класів

Робота системи розпізнавання мовців можна розбити на два основні частини, а саме, занесення користувачів в базу даних та навчання нейронної мережі та ідентифікація користувачів по висловленій пароліній фразі. На рисунку 5 представлена UML діаграму послідовності яка графічно відображає послідовність реалізації процесів в системі під час роботи.

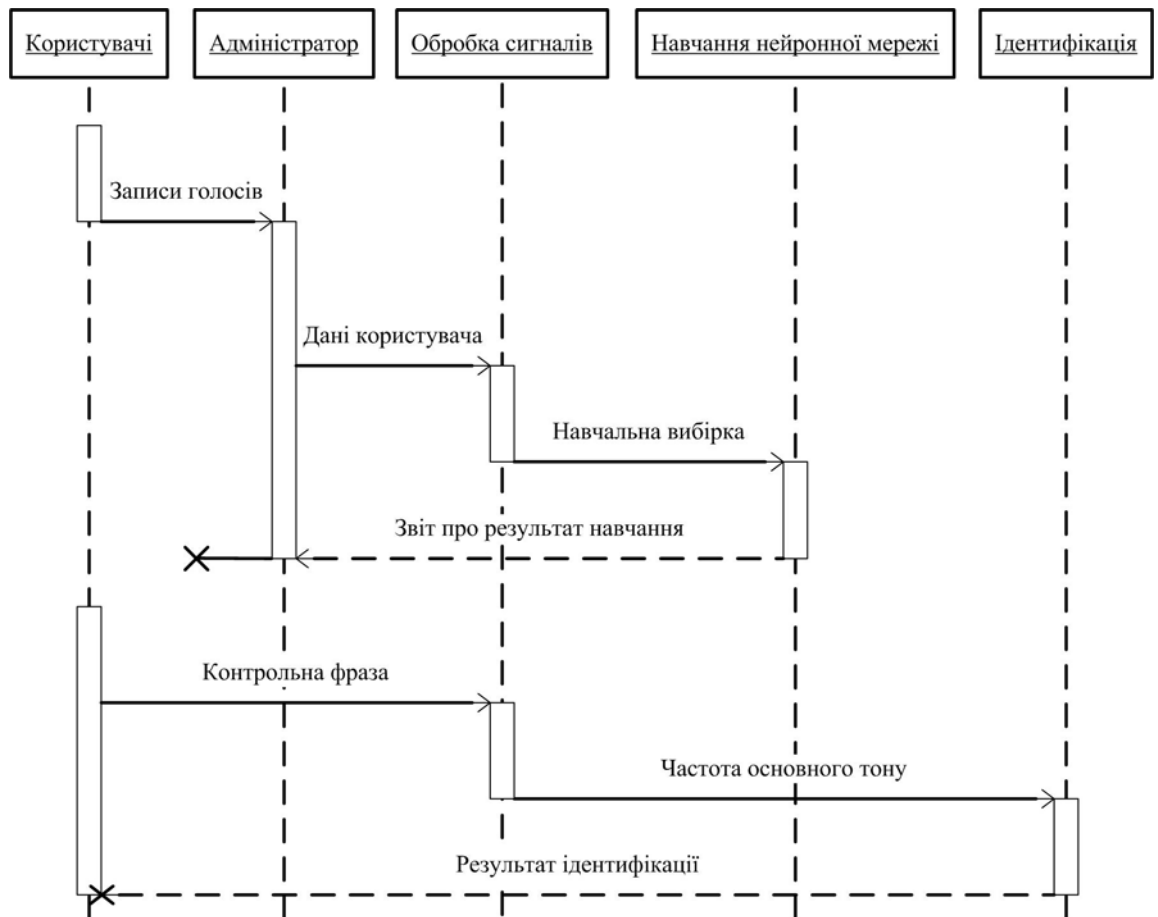


Рисунок 5 – Діаграма послідовності

Тестування системи розпізнавання мовців

До складу створеної системи увійшли, зокрема, ряд утиліт, а саме, звуковий редактор, інтерфейс створення нейронної мережі, інтерфейс навчання нейронної мережі, інтерфейс тестування нейронної мережі.

Звуковий редактор дає можливість відкривати існуючі або створювати нові звукові записи, переглядати їх основні характеристики такі як амплітуда звукового сигналу, енергія та спектр, а також дає можливість візуально побачити результат виділення частоти основного тону. Приклад роботи підпрограми представлено на рисунку 6.

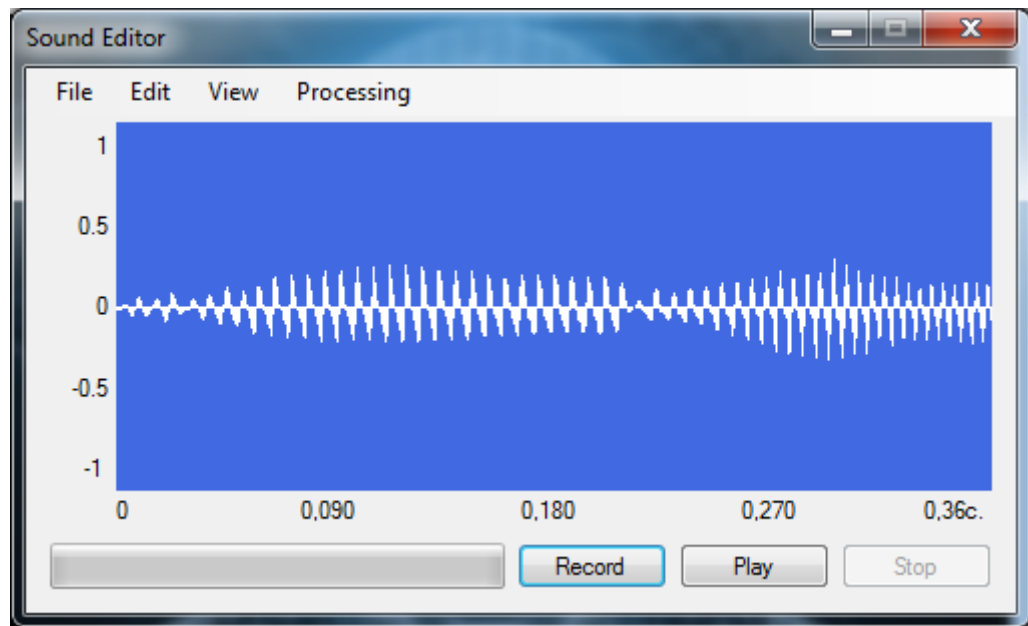


Рисунок 6 – Звуковий редактор системи

На рисунку 7 представлено результат виділення частоти основного тону синтезованої фрази «Maria» за допомогою синтезатора голосу Microsoft Anna. Як видно із рисунка, частота основного тону складає приблизно 230 – 250 Гц. Для перевірки правильності роботи алгоритму виділення основного тону, порівняємо результат із результатом програмного комплексу Colea. Результат роботи програмного комплексу Colea представлено на рисунку 8. Як видно з рисунка 7 середня частота основного тону синтезованої фрази складає 232.12 Гц. Отже алгоритм виділення частоти основного тону працює коректно.

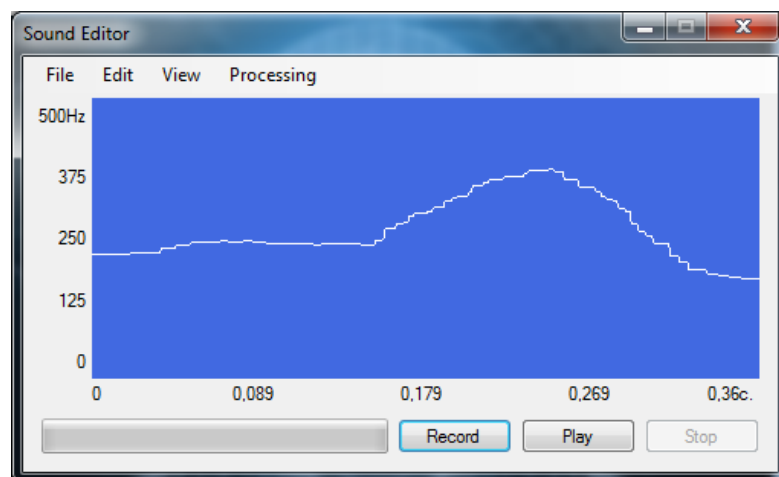


Рисунок 7 – Частота основного тону синтезованої фрази «Maria»

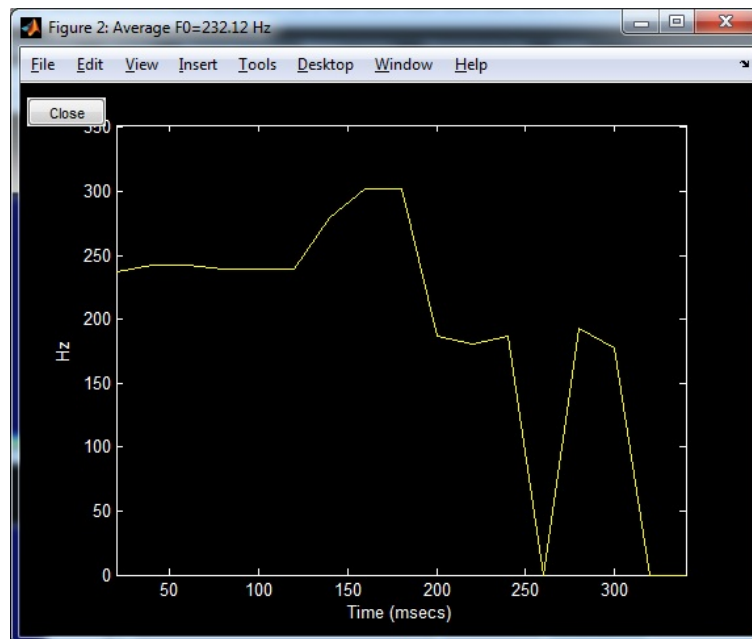


Рисунок 8 – Частота основного тону синтезованої фрази «Марія» у програмному комплексі Colea

Утиліта створення нейронної мережі реалізовано у візуальному інтерфейсі для бібліотеки алгоритмів ALGLIB, за допомогою якої можна вибрати одну із доступних архітектур мережі, вказати кількість прихованих слоїв, задати кількість нейронів у кожному слої мережі, та параметри функції активації вихідного слою якщо це потрібно. Вигляд підпрограми представлено на рисунку 9.

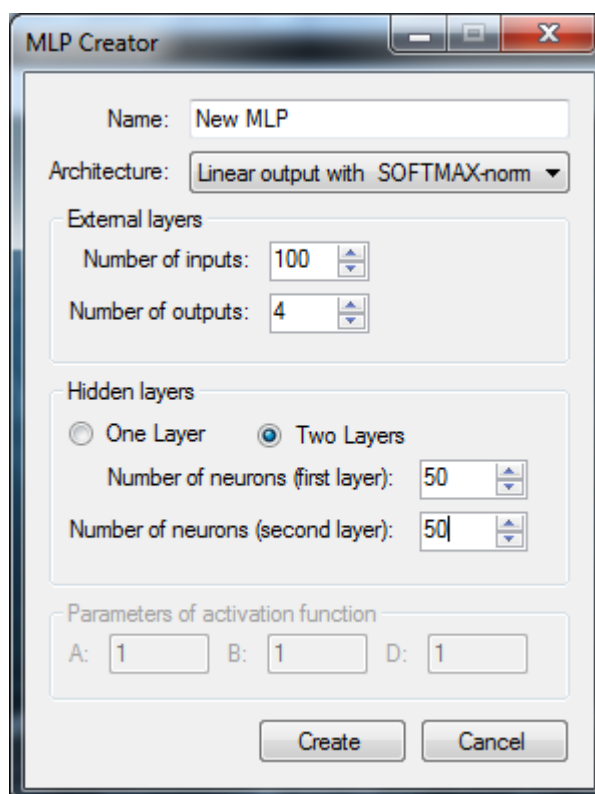


Рисунок 9 – Інтерфейс створення нейронної мережі

Також було створено утиліту навчання нейронної мережі, яка дає можливість сформувати навчальну вибірку, вибрати алгоритм по якому буде здійснюватись навчання мережі, здійснити навчання та переглянути звіт, у якому міститься інформація щодо результату навчання та кількості основних обчислень, та утиліту тестування нейронної мережі, яка дає можливість вибрати запис голосу особи, та на основі цього запису проводить ідентифікацію.

В роботі було поставлено експеримент для визначення залежності між точністю (ймовірністю) правильного розпізнавання мовців, часом на навчання системи, методом навчання нейромережевого класифікатора та кількістю нейронів на проміжному шарі пересептрона. Результати експерименту наведено у таблиці 1.

Таблиця 1 – Результати оптимізації нейромережевого класифікатора

N_2	Метод навчання	Короткочасна енергія сигналу на частоті n катних гармонік основного тону		Коефіцієнти лінійного прогнозу		Розклад Карунена – Лоева обвідної короткочасної енергії мовного сигналу	
		t, c	$T, \%$	t, c	$T, \%$	t, c	$T, \%$
10	Levenberg-Marquardt backpropagation	62	69	68	62	63	67
	Gradient dencent backpropagation	108	94	112	89	96	88
	Gradient dencent with momentum backpropagation	111	94	108	89	85	87
	RPROP backpropagation	107	95	105	90	89	88
20	Levenberg-Marquardt backpropagation	64	69	67	68	65	72
	Gradient dencent backpropagation	112	95	115	92	104	89
	Gradient dencent with momentum backpropagation	110	94	108	93	99	91
	RPROP backpropagation	110	96	116	95	97	93
30	Levenberg-Marquardt backpropagation	68	70	69	65	65	71
	Gradient dencent backpropagation	118	95	121	94	108	90
	Gradient dencent with momentum backpropagation	119	95	117	94	100	90
	RPROP backpropagation	119	96	119	96	98	93

Висновки

Авторами створено класифікатор на базі трьохшарового персептрону для системи розпізнавання мовців та проведено аналіз залежності часу навчання та точності ідентифікації від структури нейромереж та методів навчання. Виявилось, що оптимальна кількість нейронів на прихованому шарі становить 20, найкращі результати ідентифікації показали нейромережі, навчені методом RPROP backpropagation.

Найефективнішими ознаками для вирішення задачі ідентифікації диктора виявилися коефіцієнти лінійного прогнозу (96%) та значення енергії мовного сигналу на частоті основного тону диктора та на n кратних

гармоніках (96%), спектральний аналіз з додатковим врахуванням тривалості вимови наголошеної голосної літери парольного слова (94%).

Список літератури

1. Биков М.М, Ковтун В.В. Вибір інформативних ознак в задачі ідентифікації диктора // Тези доповіді на Міжнародній конференції з індуктивного моделювання, "МКІМ-2002".– Том 1.– Львів: НУЛП.– 2002.– С. 280–286.
2. Ту Д.Т, Гонсалес Р.К. Принципы распознавания образов / пер. с англ. И.Б. Гуревич; Под ред. Р.И. Журавлева.– М.: Мир, 1978.– 411 с.
3. Минский М., Пейперт С. Перцептроны.– М.: Мир, 1971.– 261 с.